5|PRTS
09/719709
528 Rec'd PCT/PTO 12 DEC 2000

A METHOD AND ARRANGEMENT CONCERNED WITH THE TRANSMISSION OF IMAGES

FIELD OF INVENTION

EL524957846US

The present invention relates to a method and to arrangement for coding and extracting regions of interest (ROI) in the transmission of still images and video images. The method and the arrangement are particularly well suited for transform-based coders, such as wavelets and DCT.

DESCRIPTION OF THE BACKGROUND ART

In transmission of digitized still images from a transmitter to a receiver, the image is usually coded in order to reduce the amount of bits required for transmitting the image.

The bit quantity is usually reduced, because the capacity of the channel used is limited. A digitized image, however, consists of a very large number of bits. When transmitting an image that consists of a very large number of bits over a channel which has limited bandwidth, transmission times will be unacceptably long for the majority of applications if it is necessary to transmit every bit of the image.

Consequently, in recent years research has been directed to coding methods and techniques for digitized images with the object of reducing the number of bits necessary to transmit the images.

These methods can be divided into two groups:

Lossless methods, i.e. methods exploiting the redundancy in the image in such manner as to enable the image to be reconstructed by the receiver without loss of information.

Lossy methods, i.e. methods that exploit the fact that not all bits are equally as important to the receiver. Hence, the image received is not identical to the original but looks sufficiently like the original image to the human eye, for instance.

In some applications, certain parts of the transmitted image are of more interest than the remainder of the image, and better visual quality of these parts of the image is therefore desired. Such a part is usually called the region of interest (ROI). Applications in which this can be useful include, for example, medical databases or the transmission of satellite images. In some cases, it is also desired, or necessary, to transmit the region of interest loss-free, while the quality of the remainder of the image is of less importance. There are also occasions when it is required to extract the regions of interest from the bit stream and decode these regions of interest without needing to decode the image as a whole.

Swedish Patent Applications SE 9703690-9 and SE 9800088-8 both describe how a mask can be calculated for delimiting such a region of interest (ROI).

SUMMARY OF THE INVENTION

The present invention addresses the aforesaid problem of defining and transmitting regions of interest and background

regions of mutually different qualities in the transmission of images.

The basic concept of the invention in solving the problem is to transform the image and to define in said transform a mask that corresponds to the regions of interest and to the background regions. The region definition and the image transform are transmitted to a receiver capable of recreating the image with the quality desired in the predetermined regions.

More specifically, the solution involves dividing the image into the desired regions. The image is then transformed to some type of transform coefficients. A mask corresponding to the separate regions in the image is defined in the transform domain and the coefficients classified and assigned to different segments in accordance with the mask definition. The segments thus belong to the corresponding regions in the image. The segments and the coefficients are transmitted in a compressed state to a receiver that is capable of reproducing regions in the image on the one hand and of reproducing the actual image on the other hand with the desired image quality in the various regions.

One advantage afforded by the invention is that several different regions of interest can be defined.

Another advantage is that different regions can have several different degrees of image quality.

Still another advantage is that only those parts of the image that are of vital interest to the user need be decoded, while avoiding decoding of the whole of the image.

Yet another advantage is that the segments can be coded independently of each other.

The invention will now be described in more detail with reference to preferred embodiments thereof and also with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block schematic illustrating an inventive arrangement.

Figure 2 is a flow chart illustrating part of an inventive method.

Figure 3 is a flow chart illustrating a further part of an inventive method.

Figure 4 is a diagram illustrating classification of transform coefficients.

Figure 5 is a diagram for interlinking image segments in a bit stream.

Figure 6 is a view of an image with object.

Figure 7 is a graphic representation of the topology in Figure 6.

DESCRIPTION OF PREFERRED EMBODIMENTS

Sub A1

Figure 1 is an overview of an arrangement for coding and transmitting images. An image 3 of an object is stored in digital form in a digital camera 1, and the image presented on a screen 4. The screen is connected to a computer 2 which is programmed to divide the image 3 into objects or regions, of which a background region R1 and regions of interest R1 and Rn are shown. An image coder 5 in the computer 2 wavelet-transforms the image, while simultaneously compressing the image, and generates a compressed bit stream PS1. An operator at the image screen 4 defines the regions of interest R2 and Rn. The image coder includes means for creating a mask PS2 in accordance with the regions and defines separate parts, segments, of the bit streams with respect to the corresponding regions R1, R2 and Rn, with the aid of said mask. The definition also enables the regions R1, R2, Rn in the form of said separate segments in the bit stream PS1 to be coded to different degrees of accuracy. A transmitter 6 sends the bit stream, including the definition of the positions and shapes of the regions R2 and Rn to a receiver 7 which is connected to a computer that includes an image decoder 8. The decoder decodes the bit stream PS1 and reproduces the mask definition PS2 and presents the image on an image display screen 9. The accuracy of the background R1 is relatively poor, whereas each of the regions R2 and Rn has respectively a higher degree of accuracy.

The following definitions are given in order to assist in describing the inventive method:

- A segment is defined here as all of the coefficients in the transform domain that belong to a given object or the background in the image. The segment can then be divided further into subsets.

- A subset is defined here as a number of coefficients in a part of the transform domain (e.g. a subband in the case of the wavelet transform) which is required for the reconstruction and which belongs to a segment in the digitized image, see Figure 4.

As before mentioned, the coefficients are classified and can be assigned to individual segments. When this classification is made, the segments are coded independently of one another to different levels of accuracy, which yields a bit stream for each segment. These segments are then joined together.

The inventive encoding method will be described with reference to Figure 2. The digitized image 3 to be transmitted presents the background R1 and the regions of interest R2 and Rn. The following procedural steps are carried out:

1. Perform a transformation of the image 3 according to step 21. In the illustrated case, this transformation is performed with a wavelet transform or with a discrete cosine transform (DCT).

Sub A2

2. Create a mask according to step 22 with the aid of information as to how the digitized image 3 shall be divided into the background R1 and the objects R2 and Rn. The techniques described in Swedish Patent Applications SE

9703690-9 and SE 9800088-8 can be used to this end. The mask is created in the transform domain and describes which coefficients are required to reconstruct the different objects or the background. Different segments SG1, SG2 and SGn correspond to the background R1 and the objects R2 and Rn.

3. Use the mask to classify the transform coefficients as belonging to the different segments SG1, SG2, SGn, according to step 13.

4. Code the segments independently of one another, according to step 24. This gives the number of bits needed for each subset.

5. Concatenate the subset streams together with necessary substream information and header information, according to step 26. This requires a bit stream description, given below.

6. Send the concatenated bit streams 27. This includes shape data 271, bit stream information 272, subband 0 referenced 273 and subband 1 referenced 274.

The method enables the receiver to have immediate access to any parts of the image when so desired, as shown in Figure 3. This is possible because the information as to where different parts are found in the bit stream is known.

One method of how the decoder may work is described below with reference to Figure 3.

1. Receive the bit stream 27 and decode the header information required, according to step 31.

2. Find and decode the required segment information, step 32.

3. Create a mask in the transform domain, for instance with the aid of the technique described in said Patent Applications SE 9703690-9 and SE 9800088-8; step 33. The mask describes those coefficients that are required to reconstruct the desired objects or background.

4. Decode requisite segment data from the bit stream; step 34.

5. Reconstruct the requisite segments; step 35.

6. Decode and show the image; step 36.

BIT STREAM DESCRIPTION

A description will now be given of those components in the bit stream 27 that are required when applying the described technique.

Data structures and pointers

Pointer

A pointer is a set of symbols that defines the position of a bit or a byte in a bit stream or a file. Many ways of defining a pointer have been defined in computer science. Any

one of these methods can be used here. A pointer can be defined implicitly by a specific bit stream composition rule. A pointer can be defined relative to an explicitly or implicitly determined position. A simple way of defining a pointer is to determine the number of bits between the requested position and a known reference point, such as the first bit in the bit stream, for instance.

Topology descriptor

The topology descriptor, TOP, is a set of symbols that defines the topological relationship between numbered objects and shapes. This is illustrated in Figure 6, in which four objects 01, 02, 03, 04 and four shapes S1, S2, S3 and S4 are shown. The topology of the image can be represented, e.g., as a tree graph as shown in Figure 7. The nodes and the edges of the tree graph can be coded in a data structure using well known methods. P_TOP is a pointer to a topology descriptor.

Shape descriptor

A shape descriptor, S_i, defines the appearance of a closed boundary line of an object. The shape number, *i*, is given by a topology descriptor. Many different shape coding techniques can be used. Examples of such methods are chain coding and shape coding methods in MPEG-4. Shape descriptors can be decoded independently of one another once their respective positions in the bit stream is known. P_S_i is a pointer to a shape descriptor.

Segment descriptor

A segment descriptor, T_i, is a compressed set of symbols that encode a segment as described above. The segment includes an ordered set of subsets. The object number, *i*, is given by a topology descriptor. p_T_i is a pointer to a segment descriptor.

Subset descriptor

A subset descriptor, B_{ij}, is an independently decodable subset, *j*, of a segment descriptor, T_i, which describes, e.g., the coefficients that belong to a given subband, *j*, as described above. p_B_{ij} is a pointer to a subset descriptor.

Multiplexed segment descriptor

Several segment descriptors, $\{T_i, T_j, T_k \ldots\}$, can be multiplexed into a common data structure MT(i,j,k). This is done normally for the purpose of simultaneous progressive transmission of a set of objects. The data structure, MT, is called a multiplexed segment descriptor. Several multiplexing methods can be used. p_MT is a pointer to a multiplexed segment descriptor.

Segment multiplexing methods

Examples of multiplexing methods are shown in Figure 5. A simple method is to interleave subsets 52 belonging to the component segments so that:

$$MT(i,j,k)=\{B_{i0}, B_{j0}, B_{k0}, B_{i1}, B_{j1}, B_{k1}, B_{i2}, B_{j2}, B_{k2}\ldots\}$$

In this case, the order of the symbols corresponds to the order in the bit stream 51, with symbols on the left being sent first. Subsets in a multiplexed stream may be excluded if they are known by the decoder.

Bit stream storage format

In order to obtain immediate access to any object whatsoever in the image, the stored bit stream or file structure should preferably include at least the following components:

In the image header, if required:

Topology descriptor TOP

Pointers to shape descriptors $\{p\_S_1, p\_S_2...p\_S_N\}$

Pointers to segment descriptors $\{p\_T_0, p\_T_1,...p\_T_N\}$

Optional pointers to subset descriptors: for each $k=[0,N]$, $\{p\_B_{k0}, p\_B_{k1},...p\_B_{kN}\}$

In the actual stored bit stream if needed:

Shape descriptors $\{S_1, S_2,...S_N\}$

Segment descriptors $\{T_0, T_1,...T_N\}$

A group of segment descriptors with index $\{k,l,m...\}$ can optionally be replaced with a multiplexed segment descriptor $MT(k,l,m...)$

N is the number of stored objects. The background is the object with index O.

PROGRESSIVE TRANSMISSION WITH IMMEDIATE ACCESS TO OPTIONAL OBJECTS

A server receives a request for sending image data to a client. The image is stored with the server in the format described in the preceding passage. Part of the stored data structures (topological data, shapes, segments and subsets) may have already been sent to the receiving terminal. This section of the description describes a procedure for composing a bit stream with the server that handles the request.

Example

Request from user

A simple request contains the following information:

Send objects with numbers k, l, m ... with a respective accuracy of n_k, n_l, n_m where the accuracy is the index for the highest subset that is sent for each index.

Several primitive requests may be sent. They will be served in the order in which they are received or in an otherwise specified order.

Procedure for serving a request (details)

Send topological information if needed. TOP is sent in response to a first request for image information.

Send all shape descriptors that are necessary to describe the boundaries of the objects requested. It is not necessary to send shape descriptors that are already known to the decoder. When using the topological tree structure in Figure 7, it is found that not all shape descriptors on the same branch as the object or on the same or lower hierarchical level need be sent. The server knows the state of the decoder and will send solely those shape descriptors that are unknown to the decoder.

Send (multiplexed) subset descriptors that describe the objects requested to the defined accuracy. Subset descriptors that are already known to the decoder need not be sent. For instance, the user is aware of the subsets $\{B_{k0}, B_{k1}, B_{k2}, B_{k3}\}$ belonging to segment k. Subset descriptors $\{B_{k5}, B_{k6}, B_{k7}\}$ must be sent when object k is requested to accuracy 7.

EXAMPLES

In this section of the description, examples are given with respect to situations in which the proposed method can be applied.

Assume, according to Figure 5, that in the centre of the image R51 there is an encircled region R52 whose quality must be better than the quality of the region R53 outside the circle, this latter region being referred to hereinafter as

the background. However, both the background R53 and the region R52 shall be transmitted simultaneously. The following then takes place:

1. The original image is transformed with a wavelet transform.

2. A mask is then created in the transform domain. This mask describes the coefficients that are required in the transform domain in order to reconstruct the region R52 and the background R53. The created mask is then used to classify the coefficients in the transform domain in two segments, one segment for the region and one segment for the background. The two segments are built up by a number of subsets. In the illustrated case, the number of subsets is the same as the number of subbands in the transform domain. The situation on hand is thus:

2.1 In respect of the region segment belonging to the region R52:

$$\{\{r_{0,1}, r_{0,2}, \ldots, r_{0,1}\}, \ldots \{r_{no\_subbands,1}, r_{no\_subbands,2}, \ldots r_{no\_subbands,j}\}\}$$

where i,j are the number of coefficients in the different subsets.

2.2 In respect of the background segment belonging to the background R53:

$$\{\{b_{0,1}, b_{0,2}, \ldots b_{0,p}\}, \ldots, \{b_{no\_subbands,1}, b_{no\_subbands,2}, \ldots b_{no\_subbands,q}\}\}$$

where p,q are the number of coefficients in the different subsets.

3. The two subsets are then coded as follows:

3.1 In respect of the region segment:

A shape descriptor $T_r=\{B_{r,0},B_{r,1},\ldots,B_{r,no\_subbands}\}$ and a set of subset pointers $\{p\_B_{r,o},p\_B_{r,1},\ldots,p\_B_{r,no\_subbands}\}$.

3.2 In respect of the background segment:

A segment descriptor $T_b=\{B_{b,0},B_{b,1},\ldots,B_{b,no\_subbands}\}$ and a set of subset pointers $\{p\_B_{b,0},B_{b,1},\ldots,p\_B_{b,no\_subbands}\}$.

4. The two segments are then combined into a single bit stream, bit stream 51, in the following manner:

<image header><TOP><S_r><$\{p\_B_{b,0},p\_B_{r,0},p\_B_{b,1},p\_B_{b,no\_subbands},$ $p\_B_{r,no\_subbands}\}$><$MT(b,r)=\{B_{b,0},B_{r,0},B_{b,1},B_{r,1},\ldots,B_{b,no\_subbands},$ $B_{r,no\_subbands}\}$>

In this case, the subsets are combined in the manner shown in the upper part of Figure 5, with the sub-bit streams 52 of the region being taken alternately with the sub-bit streams of the background. It will be noted that the TOP field is not required when the receiver is aware of the order in which the various parts of the image are set. The first part of the array, from <image header> to ...p_B...}> is, in other words, a definition of where the different image regions are placed in the remainder of the compressed bit stream <MT(b,r)={...B...}>.

5. The combined bit stream is then sent to the receiver.

The following takes place on the decoder side:

6. The image header together with the topology, shape information and pointers are read.

7. The decoder is now able to create the same mask as that described above.

8. The decoder creates the segments with the underlying subsets.

9. The decoder commences with decoding the combined bit stream and filling in the transmitted transform coefficients in the corresponding subsets.

10. An inverse transform is used.

11. The image is transmitted and reconstructed.

The aforedescribed is one way of using the proposed method. Other methods may be to combine (mix) the bit streams in another way. For instance, as shown in the bottom part of Figure 5, the region R52 may be transmitted first, followed by the background R53. Another example is one in which more than one region is found, as described with reference to Figure 6, wherewith these regions are combined in a number of different ways.

In addition to the earlier mentioned advantages, the proposed method has the added advantage of enabling shape information to be sent only when needed.